Appendix

Appendix A1 Study characteristics: Chambers, Chamberlain, Hurley, and Slavin, 2001 (quasi-experimental design)

Characteristic	Description
Study citation	Chambers, B., Chamberlain, A., Hurley, E. A., & Slavin, R. E. (2001, April). *Curiosity Corner: Enhancing preschoolers' language abilities through comprehensive reform.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
Participants	The study began with 448 low-income preschool children who ranged in age from two years, seven months to four years, eleven months.[1] At posttest, 316 children were included in the study with analysis samples ranging from 311 to 315. The three-year-olds were from private early childhood centers (N = 169) and the four-year-olds were from public preschools (N = 147). The final sample included 68% African-American children, 16% Caucasian children, and 11% Hispanic children; 49% of the sample were female. Eight preschools[2] (public and private) were assigned to the *Curiosity Corner* intervention group and eight preschools (public and private) matched on demographic characteristics were used as the comparison group.
Setting	The study took place in 16 preschools (a mix of public and private) in four high poverty, urban school districts in New Jersey.
Intervention	The intervention group children participated in *Curiosity Corner.* Information on duration, frequency, and intensity of implementation was not reported.
Comparison	The comparison group children participated in the regular early childhood curriculum at their preschool centers.
Primary outcomes and measurement	The primary outcome domains were children's oral language and cognition. The study used three subtests of a standardized test (the Mullen Scales of Early Learning, American Guidance Services Edition): expressive language, receptive language, and visual reception. The study also used the Early Childhood Environment Rating Scale-Revised (ECERS-R) to evaluate classroom quality, but the measure is not included in this WWC review because it is not relevant to the topic review (see Appendices A2.1 and A2.2 for more detailed descriptions of outcome measures).[3]
Teacher training	The program provided teachers with detailed instructions for the lessons in the Teacher's Manual, as well as the materials needed for the instructional activities. Teachers, teaching assistants, and administrators were trained in two-day initial training sessions, followed by six in-class visits by a Success for All Foundation (SFA) trainer. In addition, teachers were observed, mentored, and supported by *Curiosity Corner* coaches from the school districts. The coaches were trained by SFA staff over a two-year period as described in the intervention report. The coaches offered workshops to help teachers implement the curriculum.

1. Information on total sample size was provided by the study authors upon WWC request.
2. Information on the number of schools in each condition was provided by the study authors upon WWC request.
3. For further details about the outcomes included in the early childhood education topic review please see the Early Childhood Education Protocol.

Appendix A2.1 Outcome measures in the oral language domain

Characteristic	Description
Mullen Scales of Early Learning (MSEL) Expressive Language scale	A scale from a standardized measure of children's expressive language skills such as speaking and forming language (as cited in Chambers et al., 2001).
MSEL Receptive Language scale	A scale from a standardized measure of children's receptive language skills such as auditory organization, sequencing, and use of spatial concepts (as cited in Chambers et al., 2001).

Appendix A2.2 Outcome measures in the cognition domain

Characteristic	Description
MSEL Visual Reception scale	A scale from a standardized measure of children's cognitive ability to process visual patterns (as cited in Chambers et al., 2001).

Appendix A3.1 Summary of study findings included in the rating for the oral language domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Curiosity Corner* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Curiosity Corner* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Chambers et al., 2001 (quasi-experimental design)[8]								
MSEL Expressive Language scale	3–4 year olds	16/313	39.49 (5.48)	37.16 (5.31)	2.33	0.43	ns	+17
MSEL Receptive Language scale	3–4 year olds	16/315	37.70 (5.27)	37.63 (5.15)	0.07	0.01	ns	+1
Domain average[9] for oral language						0.22	ns	+9

ns = not statistically significant
MSEL = Mullen Scales of Early Learning

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendix A4.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were provided by the study authors upon WWC request.
3. The posttest means are covariate-adjusted means provided by the study authors upon WWC request. The study authors included age and PPVT-III scores at pretest as covariates in the analysis.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Chambers et al. (2001), a correction for clustering was needed, so the significance levels may differ from those found by the study authors.
9. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the cognition domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Curiosity Corner* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Curiosity Corner* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Chambers et al., 2001 (quasi-experimental design)[8]								
MSEL Visual Reception scale	3–4 year olds	16/311	42.19 (3.97)	42.88 (3.97)	–0.69	–0.17	ns	–7
Domain average[9] for cognition						–0.17	ns	–7

ns = not statistically significant
MSEL = Mullen Scales of Early Learning

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendix A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were provided by the study authors upon WWC request.
3. The posttest means are covariate-adjusted means provided by the study authors upon WWC request. The study authors included age and PPVT-III scores at pretest as covariates in the analysis.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Chambers et al. (2001), a correction for clustering was needed, so the significance levels may differ from those found by the study authors.
9. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Summary of subgroup findings for the oral language domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Curiosity Corner* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Curiosity Corner* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Chambers et al., 2001 (quasi-experimental design; 3 year olds)[7]								
MSEL Expressive Language scale	3 year olds	16/167	39.26 (5.04)	37.54 (4.30)	1.72	0.36	ns	+14
MSEL Receptive Language scale	3 year olds	16/168	37.76 (4.40)	37.52 (4.68)	0.24	0.05	ns	+2
Chambers et al., 2001 (quasi-experimental design; 4 year olds)[7]								
MSEL Expressive Language scale	4 year olds	12/146[8]	43.58 (4.55)	43.29 (4.01)	0.29	0.07	ns	+3
MSEL Receptive Language scale	4 year olds	12/147[8]	43.10 (4.32)	42.85 (3.75)	0.25	0.06	ns	+2

ns = not statistically significant
MSEL = Mullen Scales of Early Learning

1. This appendix presents subgroup findings for measures that fall in the oral language domain. Total group scores were used for rating purposes and are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The study authors included age and PPVT-III scores at pretest as covariates in the analysis.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Chambers et al. (2001), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
8. The sample size for the comparison group reported in the original study was incorrect and the correct sample size was provided by the study authors.

Appendix A4.2 Summary of subgroup findings for the cognition domain[1]

Outcome measure	Study sample	Sample size (schools/ children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Curiosity Corner* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Curiosity Corner* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Chambers et al., 2001 (quasi-experimental design; 3 year olds)[7]								
MSEL Visual Reception scale	3 year olds	16/165	42.32 (3.54)	42.66 (4.04)	–0.34	–0.09	ns	–4
Chambers et al., 2001 (quasi-experimental design; 4 year olds)[7]								
MSEL Visual Reception scale	4 year olds	12/146[8]	45.49 (3.20)	45.61 (3.20)	–0.12	–0.04	ns	–1

ns = not statistically significant
MSEL = Mullen Scales of Early Learning

1. This appendix presents subgroup findings for measures that fall in the cognition domain. Total group scores were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The study authors included age and PPVT-III scores at pretest as covariates in the analysis.
4. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Chambers et al. (2001), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
8. The sample size for the comparison group reported in the original study was incorrect and the correct sample size was provided by the study authors.

Appendix A5.1 *Curiosity Corner* rating for the oral language domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of oral language, the WWC rated *Curiosity Corner* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. The study did not show statistically significant or substantively important effects, either positive or negative.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study examined effects on oral language.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The study did not show statistically significant or substantively important negative effects.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. The study did not show statistically significant or substantively important positive effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Not met. The study did not show statistically significant or substantively important effects, either positive or negative.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. The study did not show statistically significant or substantively important effects, either positive or negative.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. The study did not show statistically significant or substantively important effects, either positive or negative.

(continued)

Appendix A5.1 *Curiosity Corner* rating for the oral language domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

  Not met. The study did not show statistically significant or substantively important negative effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. The study did not show statistically significant or substantively important positive effects.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study examined effects on oral language.

AND

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. The study did not show statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.2 *Curiosity Corner* rating for the cognition domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of cognition, the WWC rated *Curiosity Corner* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. The study did not show statistically significant or substantively important effects, either positive or negative.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study examined effects on cognition.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The study did not show statistically significant or substantively important negative effects.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. The study did not show statistically significant or substantively important positive effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Not met. The study did not show statistically significant or substantively important effects, either positive or negative.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. The study did not show statistically significant or substantively important effects, either positive or negative.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. The study did not show statistically significant or substantively important effects, either positive or negative.

(continued)

Appendix A5.2 *Curiosity Corner* rating for the cognition domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

  Not met. The study did not show statistically significant or substantively important negative effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. The study did not show statistically significant or substantively important positive effects.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study examined effects on cognition.

AND

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. The study did not show statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Centers	Children	
Oral language	1	16	316	Small
Print knowledge	0	0	0	na
Phonological processing	0	0	0	na
Early reading/writing	0	0	0	na
Cognition	1	16	316	Small
Math	0	0	0	na

na = not applicable/not studied

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."